**REAL-TIME SMALL BODY PRECOVERY SEARCHES IN ANY ASTRONOMICAL DATASET** Joachim Moeyens[1,2], E. Lu[1], M. Micheli[3], M. Juric[2], K. Kiker[1], A. Koumjian[1], S. Nelson[1], A. Posner[1], N. Tellis[1]; [1]Asteroid Institute, a program of B612 Foundation, 20 Sunnyside Ave, STE F, Mill Valley, CA 94941, USA; moeyensj@uw.edu; [2]DiRAC Institute & Department of Astronomy, University of Washington, 3910 15th Avenue NE, Seattle, WA 98195, USA; [3]ESA PDO NEO Coordination Centre, Largo Galileo Galilei, 1, 00044 Frascati (RM), Italy;

***Keywords:*** *precovery, characterization, risk list, cloud services, data processing*

**Highlights:**

- Automated precovery searches of risk list objects in a growing list of catalogs of observations from astronomical surveys
- Hosted on a cloud platform capable of scaling to meet the computational needs
- Final human vetting of precovery candidates before submission
- 37 observations of 19 risk list objects have been identified in three datasets so far: NOIRLab Source Catalog, Zwicky Transient Facility, and SkyMapper Southern Survey

**Introduction:**

The Asteroid Discovery, Analysis, and Mapping (ADAM) platform is a scalable cloud-based astrodynamics platform being developed by the Asteroid Institute, a program of B612 Foundation. ADAM offers a variety of astrodynamics services and standardized astronomical datasets that, together, enable compute-intensive research in planetary science and planetary defense. Newly added to the suite of tools and services is ADAM::precovery. ADAM::precovery performs at-scale searches for as-yet unattributed observations of known asteroids from point-source catalogs. Finding additional observations of objects, after their initial discovery, is important to help reduce their orbital uncertainty, thereby increasing the fidelity of their orbital propagations. ADAM's current unified data assets include observations from the NOIRLab Source Catalog [1], SkyMapper Southern Survey [2], and the Zwicky Transient Facility [3, 4, 5, 6]. For ongoing surveys such as ZTF, ADAM ingests new observations as they are made public. Work is in progress to include Pan-STARRS [7], Catalina Sky Survey [8], among others.

We describe an automated real-time application of ADAM::precovery that runs nightly as changes are detected to either the ESA and JPL risk lists and as new observations are made available in ADAM's data pipeline. We describe the different components of this service and report on the early results.

**Finding Precovery Candidates in Astronomical Survey Data:**

To perform precovery searches in data from large astronomical surveys, we developed an open-source Python package named `precovery` [9]. A critical design consideration is that searches in TB-scale data need to be both fast and parallelizable to fully leverage the power of hosting such an algorithm on a cloud-based infrastructure. To make searches fast, precovery requires data to be restructured into a "unified" data storage format. This format has two components: the first component is an index currently maintained as a database that contains information on the exposures contained within a catalog of observations. The input exposures are mapped to a common HEALPix grid [10] with a user-defined size. This grid allows multiple datasets containing observations from different telescopes to be searched simultaneously. The second component is a series of binary files that store the observations measured from those exposures. Quantities such as the observation time, the position on the sky (RA, Dec), the 1-$\sigma$ positional uncertainties, the photometric magnitude and its uncertainty, and the program-assigned observation ID are stored in these files.

The process of converting catalogs of observations into this format we call "indexing". Once indexed, `precovery` is able to perform fast searches of these data for potential precovery candidate observations. This is accomplished in a 4-step procedure per input orbit:

1. The "index" of HEALPix-mapped exposures is divided into windows of time (the default is 7 days). Once divided into windows, an input orbit is propagated with N-body dynamics to each window midpoint.
2. From each midpoint, the orbit is propagated with fast 2-body dynamics to identify any intersecting HEALPix-mapped exposures within each window[1].

[1] The chosen HEALPixel size is sufficiently large that in most cases the deviation between 2-body and N-body propagated orbits are still captured within the same pixel. This, however, does not address edge cases where the orbit is near the boundary of

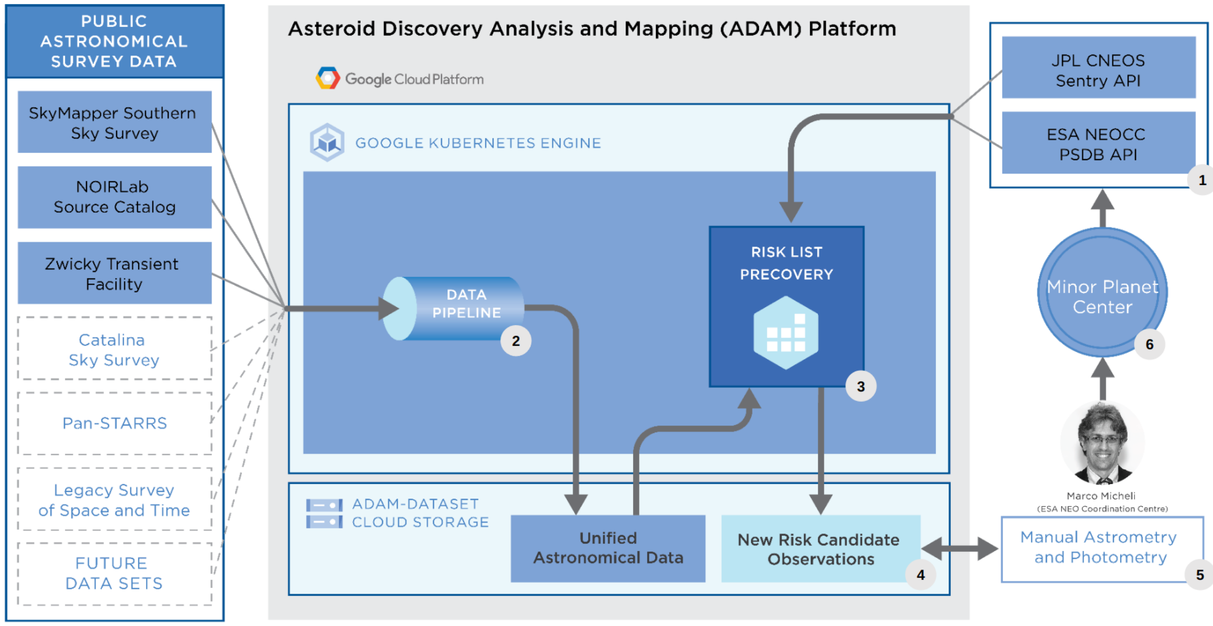


**Figure 1: ADAM's Precovery Service Applied to the Risk List. (1) Every night the ESA and JPL risk lists are automatically scanned for changes. (2) Newly available observations from ongoing surveys (such as ZTF) are automatically ingested via ADAM's data pipeline. "Indexing" occurs at the end of this step as these new data are stored into the unified data format. (3) ADAM's data assets are searched for candidate precovery observations using open-source code `precovery`. (4) A mosaic of postage stamps is generated for human validation of the candidate observations. (5) After manual vetting, the candidates are sent to Dr. Marco Micheli at the ESA PDO NEO Coordination Centre where he performs precise astrometry to take into account effects such as trailing. (6) After successful astrometric and photometric measurement, the candidates are submitted to the Minor Planet Center.**

3. Once intersecting exposures have been identified the orbit is then propagated from the window midpoint to each exposure using N-body dynamics.
4. The angular distances between the predicted position of the orbit and the observations are calculated. If any observations lie within a user-defined angular tolerance those observations are identified as possible precovery candidates.

`precovery` will output a list of potential precovery observation candidates per input orbit. If there are instances where an input orbit intersected an "indexed" exposure but no observations were found within the angular tolerance, then optionally, `precovery` can return information on the exposure. To aid in the validation of precovery results, we developed an open-source Python package called `cutouts` [11]. This software uses the metadata stored by `precovery` to query the appropriate image access service for each dataset to generate a mosaic of postage stamps centered on the predicted position of the input orbits in each image. An example of such a mosaic is depicted in Figure 2.

In Table 1, we list the current data assets maintained in ADAM.

**ADAM::precovery Applied to the Risk List:** The `precovery` and `cutouts` codes are deployed on ADAM and, in-tandem with ADAM's data assets and data pipeline, form a service we call ADAM::precovery. We now describe an application of ADAM::precovery to the ESA and JPL risk lists. The risk lists contain >1600 Near-Earth Objects (NEOs) that have a chance to impact Earth over the next few centuries. Many of these objects are initially discovered with short observation arcs and have high orbital uncertainties. It is

a pixel and will be addressed as we continue to improve the precovery code.

| Observatory Code | Date Range | Observations [millions] | Percent of Total [%] | Data Source |
|---|---|---|---|---|
| W84: DECam at CTIO | Sep 2012 - Nov 2019 | 2371 | 73.97 | NOIRLab Source Catalog (DR2) [1] |
| Q55: SkyMapper at SSO | Mar 2013 - Mar 2018 | 563 | 17.57 | SkyMapper Southern Survey (DR2) [2] |
| 695: KPNO | Jan 2016 - Aug 2017 | 186 | 5.81 | NOIRLab Source Catalog (DR2) [1] |
| V00: Bok Telescope at KPNO | Nov 2015 - Feb 2019 | 43 | 1.34 | NOIRLab Source Catalog (DR2) [1] |
| I41: ZTF at Palomar | Jun 2018 - Feb 2023 | 41* | 1.31 | Zwicky Transient Facility Alert Stream [6] |

**Table 1: ADAM's Unified Data Assets. Surveys such as ZTF perform difference imaging which removes static and non-variable sources. In cases where difference imaging is not available, we remove static sources as best as possible via filtering in our data pipeline. *Ongoing survey: observations are ingested as they are made.**

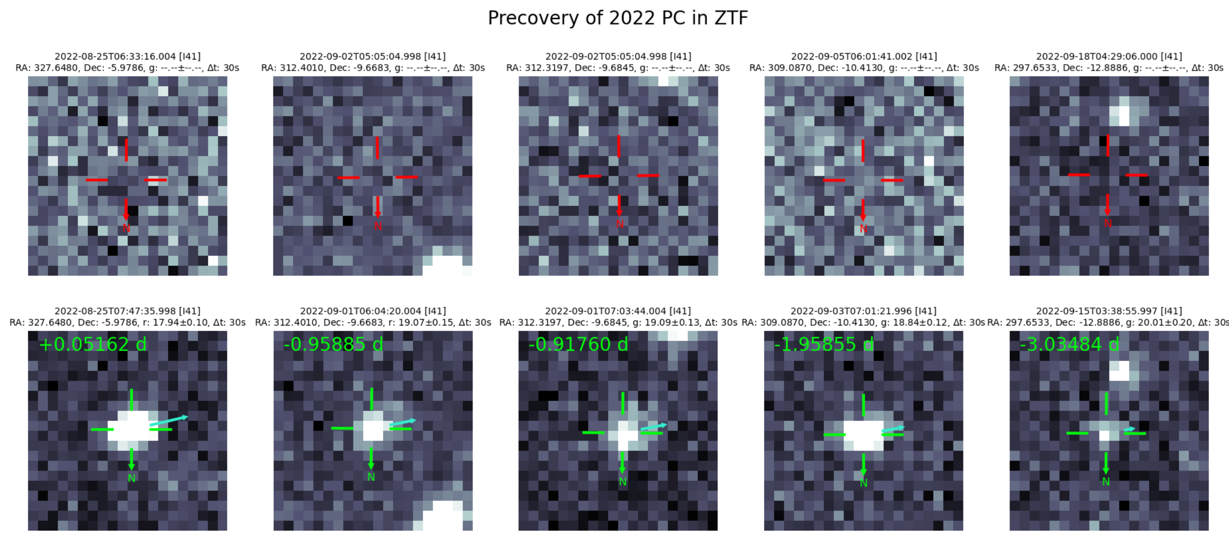


**Figure 2: A Mosaic of Postage Stamps for 2022 PC Found in ZTF Data. Each column is a postage stamp queried at the same position on the sky. The center of the crosshair represents the location queried in the image. The bottom row corresponds to images where we expect 2022 PC to appear. That is, at the location and time of the candidate precovery observations identified by ADAM::precovery. The top row represents images of the same area of the sky at a nearby time when we do not expect 2022 PC to appear. The cyan arrow represents the expected motion within the image (though offset from the center so as not to obscure the detection). Mosaics such as these offer powerful visual validation. If there consistently are sources at the center of the images where we do not expect an object to appear then that is indicative of an erroneous identification. In practice, these mosaics are generated as an animated GIF showing both the detections and comparison images in a time-ordered sequence, sometimes referred to as "blinking" cutouts. Note that the images in the bottom row appear darker due to a scaling normalization which is skewed due to the additional presence of a bright source in the image.**

therefore prudent to find additional observations of these objects after they have been discovered to decrease their orbital uncertainty and, in turn, improve the accuracy of impact probability calculations. In the best-case scenario, additional observations of these objects may remove them from the risk list altogether.

ADAM::precovery runs nightly processing of the combined ESA and JPL risk lists to aid in the effort of planetary defense. In Figure 1, we show the current pipeline for precovery searches of risk list objects. Every night the ESA and JPL risk lists are scanned for any changes and are combined into a single list with duplicate objects removed. If any new objects have been added to either list, or any existing object's orbits have been updated this triggers a precovery search (Step 1 in Figure 1). Simultaneously and in parallel with Step 1, if new observations have been made available from an ongoing survey (such as ZTF), or new astronomical datasets have been added, these observations are automatically ingested into ADAM's unified astronomical data storage format and trigger a precovery search (Step 2). When a new precovery search is triggered due to any changes in the upstream assets, ADAM::precovery will start a cloud-hosted precovery search with the precovery code described in the previous section (Step 3 in Figure 1). At the conclusion of this stage of the pipeline, ADAM::precovery produces a list of potential precovery observation candidates and optionally a list of exposures where a detection could have occurred but none was found. These steps are fully automated and run in real-time as changes to the upstream assets are detected, namely changes to the risk list objects or changes to the survey data maintained in ADAM.

The second half of the pipeline relies on human vetting of the candidate precovery observations. In addition to a list of candidate observations, ADAM::precovery will produce a mosaic of postage stamps for human validation of the identified candidate precovery observations (Step 4 in Figure 1). One such example is depicted for risk list object 2022 PC in Figure 2. Each mosaic is inspected to ensure that effects such as cosmic rays, and misidentified static sources, are excluded from any submission to the MPC. Viable candidates are selected by our team and sent for final validation to Dr. Marco Micheli at the ESA PDO NEO Coordination Centre. There, Dr. Micheli, performs precise astrometry to re-measure the sources from images to take into account effects such as trailing which are typically not accounted for in point-source catalogs (Step 5 in Figure 1). After successful astrometric and photometric measurements are extracted, observations are submitted to the Minor Planet Center (Step 6 in Figure 1).

After acceptance by the MPC, the JPL and ESA pipelines will include any new observations in their orbit fitting procedures and produce new impact probability estimates and make these available to the community (this typically occurs within a day or two of submission).

ADAM::precovery, applied to the ESA and JPL risk lists as described above, has so far led to the submission of 37 observations of 19 risk list objects in archival observations from the NOIRLab Source Catalog, SkyMapper Southern Survey, and the Zwicky Transient Facility. We stress that these numbers represent only an early demonstration of the capabilities of this service and that we expect to find many additional observations as the processing is further refined and automated, as new data is made available, and as additional objects are added to the risk list.

**Results and Future Work:** The Asteroid Institute has launched a new precovery service named ADAM::precovery that scans for changes to the ESA and JPL risk lists nightly. If any changes are detected or new observations have been made by ongoing surveys, ADAM's data assets which currently include the NOIRLab Source Catalog, SkyMapper Southern Survey, and the Zwicky Transient Facility, are searched for candidate precovery observations. All observations are vetted by our team before submission to the MPC.

ADAM::precovery thus far has found 37 observations of 19 risk list objects in these data.

Running a full-scale precovery search on all $>$1600 risk list objects across $>$840 000 exposures containing 3.2 billion observations takes $\sim$3 hours on a Google Cloud VM with 30 cores. We note that this can be easily scaled to reduce processing time.

ADAM's data assets currently include the NOIRLab Source Catalog, SkyMapper Southern Survey, and the Zwicky Transient Facility. We plan to extend the pipeline to include Pan-STARRS [7], Catalina Sky Survey [8], among others, and in the future, the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) [12]. Leveraging the innate scaling of a cloud-based system such as ADAM will be vital for LSST which is predicted to quadruple the known population of NEOs [13]. Automated real-time precovery searches of LSST discoveries across multiple datasets will be important for additional characterization and orbit refinement for those NEOs which may pose an impact risk.

As the scale of precovery searches increases so too will the amount of manual vetting of precovery candidate observations. We will be examining the possibility of automating the trailed-source fitting procedure as well as building deep-learning models for additional filtering in an effort to reduce the amount of human vetting involved. In addition to upgrades to the ADAM::precovery pipeline, we are planning to add a variety of components to the precovery code to enhance precovery searches. Namely, orbital covariance propagation and mapping, and probabilistic matching on the sky that takes into account the reported observational uncertainties. Finally, we will also be adding iterative orbit determination as a method for arc extension and further validation.

**Acknowledgements:** ADAM and this research was made possible by leadership gifts from William K. Bowes Jr. Foundation, Tito's CHEERS, McGregor Girand Charitable Endowment, P. Rawls Family Fund, Steve Jurvetson, Yishan Wong and Kimberly Algeri-Wong and three anonymous leadership donors plus donors from 46 countries worldwide. Asteroid Institute is a program of B612 Foundation. ADAM is a project of the Asteroid Institute.

**References:** [1] D. L. Nidever, et al. (2021) *The Astronomical Journal* 161(4):192 ISSN 0004-6256 doi.arXiv:2011.08868. [2] C. A. Onken, et al. (2019) *Publications of the Astronomical Society of Australia* 36:e033 ISSN 1323-3580 doi.arXiv:2008.10359. [3] M. J. Graham, et al. (2019) *Publications of the Astronomical Society of the Pacific* 131(1001):078001 ISSN 00046280 doi.arXiv:1902.01945. [4] E. C. Bellm, et al. (2019) *Publications of the Astronomical Society of the Pacific* 131(995):018002 ISSN 00046280 doi.arXiv:1902.01932. [5] F. J. Masci, et al. (2019) *Publications of the Astronomical Society of the Pacific* 131(995):018003 ISSN 00046280 doi. [6] M. T. Patterson, et al. (2019) *Publications of the Astronomical Society of the Pacific* 131(995):018001 ISSN 00046280 doi.arXiv:1902.02227. [7] K. C. Chambers, et al. (2016) arXiv:1612.05560. [8] S. Larson, et al. (2003) in *AAS/Division for Planetary Sciences Meeting Abstracts #35* vol. 35 of *AAS/Division for Planetary Sciences Meeting Abstracts* 36.04. [9] S. Nelson, et al. (2022) B612-asteroid-institute/precovery: v0.1 doi. [10] K. M. Gorski, et al. (2005) *The Astrophysical Journal* 622(2):759 ISSN 0004-637X doi.arXiv:0409513. [11] J. Moeyens, et al. (2022) B612-asteroid-institute/cutouts: v0.1 doi. [12] Ž. Ivezić, et al. (2019) *The Astrophysical Journal* 873(2):111 ISSN 0004-637X doi.arXiv:0805.2366. [13] R. L. Jones, et al. (2018) *Icarus* 303:181 ISSN 10902643 doi.arXiv:1711.10621.